心理科学进展 2018, Vol. 26, No. 1, 81–97
Advances in Psychological Science DOI: 10.3724/SP.J.1042.2018.00081

场景主旨加工及其机制*

李梦琪[1] 陈志敏[1] 郑元杰[2] 任衍具[1]

([1]山东师范大学心理学院; [2]山东师范大学信息科学与工程学院, 济南 250358)

摘 要 场景主旨是指观察者在一次注视场景的过程中所获得知觉和语义信息。近年来, 场景主旨加工研究已经成为视知觉领域的重要内容, 对该问题的研究将有助于揭示视觉信息加工的机制, 对智能机器视觉的研制也有一定的借鉴意义。对场景主旨加工的影响因素、争议性的问题以及场景主旨的神经基础进行评论; 未来可以在场景主旨加工的基本单元、相关的理论解释、层级加工的调节因素、注意的调节作用、时间动力特性和脑功能网络的构建等方面做进一步的探讨。

关键词 场景主旨; 层级加工; 注意
分类号 B842

1 引言

人类的生存离不开对周围环境的感知与判断, 因此人类的视觉系统进化出了可以在极短的时间内获取场景中的必要信息, 并进一步对其做出类别判断的能力(例如: 这是一幅街道的景象)。场景主旨(scene gist, 也称 scene schema)被界定为观察者在单次注视(数百毫秒)场景的过程中所获得的知觉和语义表征(Fei-Fei, Iyer, Koch, & Perona, 2007; Friedman, 1979; 程昊, 2010; 见综述 Oliva, 2005; Oliva & Torralba, 2006)。例如, 我们可以将一幅场景描述为"海滩"、"卧室"或者"街道"。本文中涉及的场景主旨加工研究主要指观察者在较短时间内对场景刺激进行检测、再认或分类的研究。

场景主旨加工研究可以追溯到上个世纪六七十年代。Potter 和 Biederman 这两位心理学家及同事进行了开创性的工作, 他们发现人类视觉系统能够迅速提取场景的主旨信息并用于后续的类别判断(Potter, 1975; Potter & Levy, 1969)、目标搜索和记忆等(Biederman, 1972; 也见 Torralba, Oliva, Castelhano, & Henderson, 2006)。迄今为止, 尽管已有大量研究证实了人类视觉系统具有加工场景主旨的超凡能力, 但关于这种加工能力的认知与神经机制仍处于探索之中。已有关于场景主旨加工的研究主要涉及以下五个方面的内容: (1)场景主旨加工的影响因素有哪些? (2)场景主旨的层级加工优先性问题, 即在场景主旨加工过程中, 是上级水平("自然场景"和"人工场景"、"室内"和"室外")场景主旨, 还是基本水平("森林"、"山川"、"街道"和"建筑"等)场景主旨优先得到加工? (3)场景主旨加工的过程是否需要注意资源的参与? (4)场景主旨加工有哪些重要的理论观点? (5)场景主旨加工的神经生理基础, 即有哪些脑区参与了场景主旨的加工, 它们是如何协同作用完成场景主旨的加工?

场景主旨加工是场景知觉研究的重要内容, 有其重要的理论意义和应用价值。一方面有助于增进我们对视觉系统加工机制的理解(如, Malcolm, Groen, & Baker, 2016); 另一方面相关研究成果在机器视觉(如, Wei, Phung, & Bouzerdoum, 2016)、广告设计(如, Wedel & Pieters, 2015)、安全检查(如, Biggs & Mitroff, 2015)和医学影像诊断(如, Evans, Haygood, Cooper, Culpan, & Wolfe, 2016)等方面均得到了广泛的应用。

2 真实场景主旨加工的影响因素

场景中的哪些特征会影响场景主旨的加工

收稿日期: 2017-02-28
* 山东省自然科学基金面上项目(ZR2017MC058); 山东省高等学校人文社会科学研究计划项目(J13WH07); 山东省"泰山学者海外特聘专家"项目(TSHW201502038); 山东省高校人类认知与行为发展重点实验室; 山东师范大学"二层次"人才资助项目。
通信作者: 任衍具, E-mail: renyanju@gmail.com

呢？前人对该问题进行了大量的研究，下面将重点介绍颜色、空间频率、视野区域等因素在场景主旨加工中的作用。

2.1 颜色

Gegenfurtner 和 Rieger (2000)采用快速呈现–掩蔽–再认范式，操纵场景图片呈现与再认时的颜色信息，探讨颜色在场景编码和提取加工中的作用，结果发现，颜色信息既能够为早期感觉加工的编码提供线索，也有助于巩固场景图片的记忆表征(但见 Yao & Einhäuser, 2008)。Goffaux 等人(2005)选用 4 类将颜色作为诊断性特征的场景(沙漠、森林、峡谷、海岸)，经过处理获得 3 种颜色类型的场景：正常颜色场景、灰色场景和异常颜色场景(将原场景中的颜色进行红绿置换和蓝黄置换)；要求被试完成快速场景分类的 go/no-go 任务，结果发现视觉系统对正常颜色场景的分类最快最准确，其次是对灰色场景的分类成绩，而对异常颜色场景的分类成绩最差。Castelhano 和 Henderson (2008)采用情境偏向范式(contextual bias paradigm)探讨颜色在场景主旨加工中的作用，先向被试呈现场景图片，接着呈现掩蔽图片，掩蔽图片过后呈现物体标签(单词)，要求被试判断该标签所对应的物体与场景情境是否一致。其实验逻辑是：当场景图片的呈现时间足够长时，视觉系统能够较为充分地提取场景的主旨信息，被试对物体与场景情境是否一致的判断会更准确，在二者一致的情况下，就会出现“Yes”比“No”多的反应偏向效应，因此一旦产生了此类反应偏向效应，则意味着场景主旨在呈现的时间内得到了较为充分的激活。结果发现，相同反应偏向情况下，正常颜色场景所需要的呈现时间明显短于黑白场景所需要的呈现时间。由此可见，正常颜色场景的主旨较黑白场景的主旨激活需要的时间更少，即正常颜色信息对场景主旨的加工具有显著的促进作用。

然而，也有研究者认为颜色信息并不是场景主旨加工中的关键因素，它所产生的影响会受到其他因素的调节(如，Marx, Hansen-Goos, Thrun, & Einhäuser, 2014; Otsuka & Kawaguchi, 2009)。首先，颜色信息的诊断性会影响其在场景主旨加工中的作用。例如，在“森林”场景中，绿色具有较高的诊断性，因此绿色对场景“森林”的主旨加工具有促进作用；而在”城市”场景中，不存在一种固定的、具有诊断性的颜色信息，那么颜色信息对“城市”场景的主旨加工则没有明显的影响(Oliva & Schyns, 2000; Rousselet, Joubert, & Fabre-Thorpe, 2005)。其次，场景图像本身是否清晰也会影响场景早期阶段加工对颜色信息的利用程度。有研究发现，颜色信息有利于模糊广告的内容识别，但当广告图片变得清晰时，彩色广告的优势就不复存在了(Wedel & Pieters, 2015)。虽然目前颜色信息在场景主旨中作用存在灵活性的观点是肯定的，但其他因素所产生的调节作用的机制还需要进一步的研究。

2.2 空间频率

人类的视觉系统包含多个对不同空间频率信息敏感的通道，不同空间频率信息在面孔、物体和场景分类中有着不同的作用(Morrison & Schyns, 2001)。Schyns 和 Oliva (1994)为了探讨视觉系统在分类场景图片时对不同空间频率信息的选择偏向，要求被试完成对叠加图片 (hybrid image，将某一场景图片的低频成分与另一场景图片的高频成分叠加成一个新的图片)的分类任务，结果发现当叠加图片的呈现时间较短(30 ms)时，被试更偏向于使用其低频信息进行分类，而当呈现时间较长时(150 ms)时，被试更偏向于使用其高频信息来分类；他们由此提出场景主旨的早期加工采用的是“由粗糙到精细(coarse to fine, CtF)”的加工模式。近期的发展性研究表明 7~8 个月的婴儿(Otsuka, Ichikawa, Kanazawa, Yamaguchi, & Spehar, 2014)和 18~22 岁的年轻人(Musel, Chauvin, Guyader, Chokron, & Peyrin, 2012)均存在这种加工模式，而且这种加工也存在于场景选择性加工的大脑区域中(Awasthi, Sowman, Friedman, & Williams, 2013; Musel et al., 2014)。

需要注意的是，场景主旨的 CtF 加工模式并不是固定的，会受到其他因素的调节。首先，任务与场景材料也能影响视觉系统对空间频率信息加工的偏向性(Oliva & Schyns, 1997)。例如，当使用不同带宽的情境信息来引导被试进行场景中的物体进行快速检测时，物体的细节信息能够促使视觉系统利用高频情境信息来完成任务(Patai, Buckley, & Nobre, 2013)；叠加图片中与任务无关的信息会影响视觉系统对任务相关场景信息空间频率的反应偏向性(Rotshtein, Schofield, Funes, & Humphreys, 2010)。其次，场景分类过程中视觉系

统对空间频率信息的早期加工会受到注意模式的影响(如, Vanmarcke & Wagemans, 2016)。近期, Brand 和 Johnson (2014)采用 Navon 任务操纵被试的注意模式(注意整体/注意局部), 并结合场景分类任务探讨视觉系统对空间频率信息加工的偏向性是否会受到注意模式的调节, 结果发现整体性 Navon 任务的启动能更快将叠加场景归类为其低频信息所属的类别, 当低频信息受到抑制时, 分类速度有所减缓。总而言之, 空间频率对场景主旨加工的影响具有一定的灵活性, 但这种灵活性仅表现为对粗糙到精细加工过程的促进或干扰, 但还不足以打破这种模式的时间发展顺序。

尽管视觉系统对真实场景中不同空间频率信息的加工有其相对固定的选择偏向, 但大脑对不同空间频率信息的整合是非常迅速(在 100 ms 之内即可完成), 且整合过程几乎不需要注意的参与(Kihara & Takeda, 2010, 2012)。神经科学的研究提出了快速“M”假设, 该假设认为, 大脑皮层中通过大细胞通道传输信息较小细胞通道更快速, 在场景主旨加工过程中, 低频信息通过大细胞通道首先被相关脑区获取并形成初级表征(Hagmann & Potter, 2016), 为随后小细胞通道获得其他细节信息提供反馈和背景框架, 启动自上而下的促进作用(Kveraga, Boshyan, & Bar, 2007; Maguire & Howe, 2016; Mu & Li, 2013; 也见 Malcolm, Nuthmann, & Schyns, 2014; 但见 Potter, Wyble, Hagmann, & McCourt, 2014), 也可用于后期高频信息的加工(Kauffmann, Chauvin, Pichat, & Peyrin, 2015; Kauffmann, Ramanoël, & Peyrin, 2014)。快速”M”假设为 CtF 的观点提供了神经基础, 但目前仍缺少有力证据为相关脑区的协作运行方式做出解释, 后续还需进一步的实验验证。

2.3　视野区域

人眼视野根据投影从中心到边缘可以分为中央凹区, 副中央凹区, 外周区域, 三个区域对外部信息的分辨力不同, 中央凹区域的分辨力最高, 其次是副中央凹区域, 外周区域的分辨力最低。这种差异导致了不同视野区域在场景主旨识别上的作用也不尽相同(见评论 Loschky, Nuthmann, Fortenbaugh, & Levi, 2017)。值得一提的是知名的视觉研究期刊 Journal of Vision 在 2016 年第 2 期专辑论述不同视野区域在场景知觉中的作用(http://jov.arvojournals.org/issues.aspx#issueid=934904)。由此可见, 视野区域是近年来研究者所关注的一个重要影响因素。

最初的研究发现视觉系统仅通过外周视野所获取的低分辨率信息即可完成对场景主旨的判断。如, Larson 和 Loschky (2009)采用”窗口(window, 只保留场景的中央视野信息)”和”盲点(scotoma, 只保留场景的边缘视野信息)”范式, 要求被试完成场景−单词匹配任务, 结果发现利用边缘视野比中心视野完成匹配任务的正确率更高, 且仅利用边缘视野与利用完整视野的正确率无显著差异。Boucart, Moroni, Thibaut, Szaffarczyk 和 Greene (2013)发现观察者在场景与视野中央离心率 70°的情况下对场景主旨进行分类时, 可以达到 70%左右的正确率。上述结论在 Wang 和 Cottrell (2016)的建模研究中也得到了证实。这些证据均表明外周视野利用低分辨率信息能够完成场景分类任务。

Larson 和 Loschky (2009)认为这可能是由于外周视野较中心视野的面积更大, 二者在同一时间内获取的信息量不同所致。当控制二者的面积相同时, 表现出中央视野的加工优势。利用老年黄斑病变(neovascular age related macular degeneration, AMD)的病人进行的研究也得到了类似的结果, 该病症会引起中央视野的缺失, 结果发现无论场景呈现在视野中心还是外周, AMD 病人较正常人对主旨分类的灵敏度和反应时均降低, 且这种降低在场景刺激出现在视野中央时更严重(Thibaut, Tran, Szaffarczyk, & Boucart, 2014)。这表明中央视野在场景主旨加工中扮演更为重要的作用。

近期, Larson, Freeman, Ringer 和 Loschky (2014)采用类似的窗口和盲点范式, 对场景主旨早期加工中不同视野区域作用的时空动态进程做了探究, 通过控制目标场景后掩蔽出现的时间长度来操纵场景的加工时间。结果发现在当场景的加工时间为 24 ms 时, 中央视野的信息对基本水平场景分类任务正确率更高; 而增加到 70 ms 时, 中央视野与外周视野对基本水平场景分类任务的正确率没有显著差异。表明在早期场景主旨加工的时间进程中, 注意首先获取中央视野的场景信息, 随后注意从中心视野扩展至边缘视野以提取更多的信息。

上述的研究结果证实, 虽然视觉系统仅仅依

据外周视野的信息即可进行场景分类，但中心视野在场景主旨的加工中仍具有比外周视野更高的效率。有意思的是，只需外周视野足以完成场景主旨分类任务的观点，也暗示了场景主旨分类不需要集中的注意资源即可完成，这与注意在场景主旨加工中的作用的部分观点(见本文第三部分的相关内容)相符。

2.4 其他因素

除颜色信息、空间频率和视野区域外，场景主旨加工还会受到场景本身的边界(edge-based) (如, Fu et al., 2016; Walther & Shen, 2014)、振幅谱(amplitude spectra) (如, Hansen & Loschky, 2013; Joubert, Rousselet, Fabre-Thorpe, & Fize, 2009)、后向掩蔽(backward mask)的类型(Freeman, Loschky, & Hansen, 2015; Loschky, Hansen, Sethi, & Pydimarri, 2010)、情绪信息(Subramanian, Shankar, Sebe, & Melcher, 2014; 李毕琴，郭毕鹏，胡竹菁，罗跃嘉，2015)、观察者的观察视角(viewpoint) (Loschky, Ringer, Ellis, & Hansen, 2015)、工作记忆负载(孙琪，任衍具，傅根跃，2015)、个体差异(Vanmarcke & Wagemans, 2015; Vanmarcke et al., 2016)和先前经验/期望(Duh & Wang, 2014; Greene, Botros, Beck, & Fei-Fei, 2015; 孙雨生，张智君，吴彬星，2017)等因素的影响。现实场景中包含的信息错综复杂，任何条件的改变都有可能影响人眼对当前场景的感知。视觉系统无法在一瞥的时间内读入所有信息，但大脑能够灵活地运用自身的有限资源来获取最有诊断性并易于提取的信息，对场景主旨做出识别。

3 相关的争议性问题

3.1 场景主旨加工的层级优先性

近年来，场景主旨加工的层级加工优先性成为场景知觉研究领域的热点问题。对该问题的研究源于早期认知心理学关于概念表征通达的问题，存在两种截然相反的观点：基本水平加工优先(basic level superority)和上级水平加工优先(superordinate level superority)的观点。对同一幅场景图片，我们能够使用不同层级的概念来对其进行命名。例如，一幅森林的场景图片可以被命名为“森林”、“户外”或者“自然”，这几个概念虽然都能够在一定程度上反映森林的意义，却具有不同的内涵和外延。基本水平概念曾被认为是优先获取的场景主旨概念，如“森林”、“湖泊”、“卧室”、“厨房”等属于基本水平概念。而上级水平概念较基本水平概念有更大的外延，常用于该问题研究的上级水平场景概念是“自然”与“人工”或者“室内”与“室外”等。那么在场景主旨加工过程中，视觉系统会优先加工哪个层级的视觉信息呢？

物体加工的研究发现，人类对表征物体的基本水平概念的通达先于对上级水平概念的通达(如, Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976; 但见 Mack & Palmeri, 2015; Wu, Crouzet, Thorpe, & Fabre-Thorpe, 2015)。后来这一观点被引申到场景主旨加工的问题上。Tversky 和 Hemenway (1983)最先对场景主旨的层级加工进行研究，结果发现被试偏向于使用具有基本概念属性的词语来对场景进行描述，并由此认为场景主旨的基本水平得到优先加工。

然而，近年来研究者指出列属性的方式来检验层级加工优先性会受到属性词词频的影响，而基本水平属性词的词频更高，这可能会抵消原有的上级水平优势(见综述, Fabre-Thorpe, 2011)。该观点也得到了相关研究的支持：首先，视觉系统对上级水平信息的获取用时更短。例如，Fabre-Thorpe 领导的研究小组采用 go/no-go 范式要求被试尽可能快地对场景进行上级水平(“自然场景”或“人工场景”)或基本水平(“海洋”、“山脉”、“城市”、“街道”等)的分类判断，结果发现上级水平判断任务的耗时(380~390 ms)显著小于基本水平(400~460 ms) (如, Joubert, Rousselet, Fize, & Fabre-Thorpe, 2007; Macé, Joubert, Nespoulous, & Fabre-Thorpe, 2009)。Greene 和 Oliva (2009a, 2009b)采用心理物理学方法将场景分类任务的正确率达到 75%时所需要的场景刺激呈现时间定义为完成该分类任务所需要的最小时间阈限，将被试完成基本水平和上级水平分类任务的最小时间阈限相比较，发现上级水平阈限显著小于基本水平。其次，视觉系统对上级水平的信息获取更敏感。Loschky 和 Larson (2010)采用迫选任务，在目标场景后呈现掩蔽图片和提示词，要求被试对场景图片和提示词做类别匹配判断，结果发现在目标场景的呈现时间小于 72 ms 的情况下被试对上级水平概念的敏感度更高。近期 Sun, Ren, Zheng, Sun 和 Zheng (2016)采用双任务范式结合工作记忆任务和场景主旨类别辨别任务，探讨场景主旨

加工的层级性, 结果发现了上级水平的加工优势, 且在场景主旨类别辨别过程中, 先利用的是场景中的空间信息, 而后利用场景中的客体信息。这些研究均支持了上级水平的概念相较于基本水平能够更快被获取的观点。

然而, 常用来作为场景主旨研究的不同上级水平概念(室内−室外与人工−自然)之间的本质也并不相同。Kadar 和 Ben-Shahar (2012)将场景类别扩大到 15 种, 以考察上级水平分类优势是否具有普遍性, 结果发现在类别判断任务中, 场景主旨加工首先进行的是自然性/非自然性的判断, 随后才会进行室内/室外或是基本水平类别判断。除此之外, 室内与室外场景还具有相似的光谱特性, 这与自然和人工概念光谱特性的区别不同(Oliva & Torralba, 2001), 且采用 go/no-go 范式获得的自然/人工概念的上级水平优势在室内/室外类别中消失不见了(Banno & Saiki, 2015)。因此, 在考虑场景主旨层级加工优先性问题的过程中, 对不同上级水平概念的区别也是有必要的。

针对这两种相互矛盾的观点, 目前的研究更偏向于认为场景主旨的研究具有上级水平加工优势, 但这种优势效应并不稳定。首先, 场景的类间/类内关系会对场景主旨分类任务产生影响。Greene 和 Fei-Fei (2014)采用 stroop 范式之变式对视觉分类的自动性进行研究, 结果发现基本水平的分类是自动的, 而上级水平的分类不是, 支持基本水平优先加工的观点。其次, 基本水平类别的相似性程度(例如, 街道和市中心的相似性大于高速公路和市中心的相似性)能够调节场景主旨加工的上级水平优势, 甚至使结果反转而产生基本水平优势(Banno & Saiki, 2015)。近期的计算模型研究暗示, 场景主旨层级加工的优先顺序与刺激的相似性有关(Serre, 2016; Sofer, Crouzet, & Serre, 2015)。由此可见, 场景主旨的上级水平加工优势可能受个体词汇结构、上级水平概念种类、任务中的干扰项和不同类别间相似性等因素的影响, 但对于这种不稳定性产生的原因仍不清楚, 有待进一步的研究。

3.2　场景主旨加工对注意资源的需求

与层级加工密切相关的另一个问题就是场景主旨加工对注意资源的需求。早期的研究发现场景主旨加工是一项非常高效的认知活动, 可以在极短的时间内完成; 那么场景主旨加工是否需要注意资源的参与吸引了诸多研究者的兴趣。前人对此问题的回答正在经历一个富有争议的过程。部分研究者认为场景主旨的加工不需要注意资源的参与, 是一个自动化的过程(如, Li, VanRullen, Koch, & Perona, 2002); 另一些研究者认为场景主旨的提取需要注意参与, 注意资源的不足会导致场景主旨加工绩效的下降(如, Cohen, Alvarez, & Nakayama, 2011)。目前对该问题的研究多采用双任务范式, 即要求被试在同一时间内完成两种任务, 中心任务需要注意的参与, 考察被试集中注意来完成中心任务是否对场景主旨的分类或识别成绩产生影响, 若不产生影响, 则说明两者在注意资源的使用上不存在相互干扰, 即场景主旨的加工不需要注意的参与。

一些研究者采用字母辨别任务(Li et al., 2002; Poncet, Reddy, & Fabre-Thorpe, 2012)、元音字母有无判断任务(Walker, Stafford, & Davis, 2008)作为中心任务时发现, 视觉系统对自然场景进行识别和分类能力没有受到损害, 这意味着场景主旨加工不需要注意资源的参与。有研究者采用负启动范式获得了类似的发现(Otsuka & Kawaguchi, 2007)。还有研究者发现, 人类视觉系统可以在无意识的条件下习得场景类别规则的统计信息(Brady & Oliva, 2008)。然而, Cohen 等人(2011)对上述的研究设计提出了质疑, 认为可能是场景主旨加工需要的注意资源相对较少, 而双任务范式中用来占用注意资源的无关任务难度较小, 因此单/双任务对场景分类任务的成绩几乎不产生影响。他们采用多物体追踪任务(multiple object task, MOT)作为分散注意的中心任务, 该任务的特点是需要连续不间断的持续注意才能完成。结果发现场景主旨任务成绩显著下降, 意味着场景主旨的加工需要注意资源的参与。Mack 和 Clarke (2012)在外周采用十字架水平−竖直长短判断任务, 而中心呈现一张场景或马赛克掩蔽图片, 实验中仅仅要求被试完成十字架长短判断任务, 结果发现, 那些在任务结束后报告注意到场景图片的被试, 在十字架长短判断任务上的成绩显著低于那些没有注意到其他刺激的被试。结果表明对场景刺激的无意识注意会降低其他耗费注意资源任务的成绩(也见 Clarke & Mack, 2014)。这些结果均支持场景主旨提取需要注意参与的观点。

除此之外, Greene 和 Fei-Fei (2014)年采用

stroop 范式之变式，在场景图片中央嵌入与场景主旨类别一致或不一致的单词，并要求被试迅速对单词进行识别，结果发现当背景场景与目标单词意义一致时，对单词的识别需要时间更短，说明场景与单词的语义一致性对单词的识别产生了促进作用。该结果表明背景场景在没有任务要求的情况下得到自动化加工，暗示这种自动化并未占用注意资源。对此，Gronau 和 Izoutcheev (2017)发现当场景主旨识别作为无关任务(处于边缘视野)时，视觉系统在完成相关任务(处于中心视野)的同时，不能够自动地对场景主旨进行识别。根据 Gronau 和 Izoutcheev (2017)的结论可以认为 Greene 和 Fei-Fei (2014)研究结果产生的原因可能是视觉系统在完成中心任务的同时会对呈现在注意中心的背景进行自动化加工，这种加工建立在中心任务与无关背景重叠的空间关系之上，当两者分离且场景刺激与任务无关时，这种自动化加工就不复存在了。这一猜测还有待进一步的实验验证。

为了解释行为实验的这些相互矛盾的观点，研究者在近期的 ERPs 研究中，仍然采用双任务范式，发现在刺激呈现的 250ms 以内(大约 220 ms)，对场景刺激做自然/人工分类时的脑电变化不受注意资源是否减少的影响，但注意资源的减少会调节不同类别场景在脑电上差异表现的时间。对此研究者认为，注意资源的多少不影响场景的早期加工，但会对后期场景局部和细节信息的加工产生影响，并决定大脑对场景记忆的深度(Groen, Ghebreab, Lamme, & Scholte, 2016; Harel, Groen, Kravitz, Deouell, & Baker, 2016)。

场景主旨加工可能需要少量注意资源的参与。如果当前任务需要将注意集中在与场景存在空间重叠的刺激上，抑或当前任务难度不足以耗尽所有的注意资源，这些情况下场景主旨加工对注意资源的占用可能难以察觉；反之则不然。虽然有研究为场景主旨加工不需要注意参与提供了证据，但该研究中的不同任务相关刺激是同时呈现，这可能会对结果产生影响(Gronau & Izoutcheev, 2017)，这样的猜测还有待将来进一步的实验验证。

4 真实场景主旨的加工机制

4.1 早期的场景主旨加工理论

最初的场景主旨加工理论是以物体为中心的(object-centered)场景加工理论。该理论认为场景中的具有诊断性意义的物体是视觉系统进行场景主旨识别和分类的依据，视觉系统通过结合场景中的物体以及先前经验中物体可能出现的位置来对当前的场景主旨进行判断(如，De Graef, Christaens & D'Ydewalle, 1990)。或者可以说，视觉系统对场景中的一个或者几个突出物体的识别即可以实现对场景主旨的有效识别(Friedman, 1979)。这类理论很早就遭到质疑和反驳，因为有研究者发现观察者不需要对场景中的物体进行识别就能够对场景做出正确的分类判断。

4.2 空间包裹理论

针对以物体为中心的场景加工理论的不足，研究者又提出了以场景为中心的(scene-centered)场景加工理论(如, Schyns & Oliva, 1994)。其基本观点为，场景的整体特性才是场景主旨加工所需要的信息成分。同时，大量研究也表明，多种场景整体特性(如场景的纹理、颜色、体积或者空间频率等)的变化均会对视觉系统的场景识别过程产生影响。然而这些研究仅仅证明多种整体特征在场景加工中的作用和价值，却难以解释信息是以怎样的形式被提取和进一步加工的。

为了回答上述的问题，Oliva 及其同事提出了以场景为中心的加工理论−空间包裹(spatial envelope)理论，该理论试图寻找视觉系统在场景早期加工过程中提取的基本单元(primary element)，假设大脑仅通过对基本单元的编码和表征就能完成对场景主旨的加工。空间包裹理论提出 5 种全局特征，分别是自然性(naturalness)、开放性(openness)、粗糙度(roughness)、延伸性(expansion)、坚固性(ruggedness) (Oliva & Torralba, 2001)，并在随后的研究中将其扩展到 7 种(Greene & Oliva, 2009b)。Oliva 等人认为这些全局特征是由多种低水平特征整合形成，涵盖了不同频率的空间信息，能够为场景主旨的识别提供粗略但充分的信息资源。同时也有研究指出，全局特性的加工不受视觉疲劳的影响，被认为是一种自动化的信息加工(Csathó, van der Linden, & Gács, 2015)。

Greene 和 Oliva (2006)发现，在对场景图片进行快速分类时，对与目标场景拥有某种相同全局特性的干扰场景更容易虚报。例如，要求被试判断快速呈现的场景图片是否属于“森林”，那么与森林一样具有“低开放性”全局特征的非森林场景

较其他不具有此特征的干扰项更容易被识别为“属于森林”, 他们认为产生这种混淆的原因可能是全局特征被视觉系统用来作为主旨识别的依据。Greene 和 Oliva (2009a)利用计算机编写了以 7 种全局特征为依据的场景识别算法, 同时使用该计算机算法和人类被试来完成相同的场景分类任务, 结果发现二者在反应时、正确率甚至虚报率上均无显著差异, 表明计算机全局特征算法较好地模拟了人类视觉系统在场景识别和分类中的决策偏向, 支持了场景主旨能够仅仅通过对全局特征提取和整合而获得的理论假设。然而, 最近的一项研究采用重复盲视(repetition blindness)范式发现, 观察者在加工场景的前 100~150 ms 内, 所表征的是局部的视觉特征, 而不是更抽象的类别特征(Goldzieher, Andrews, & Harris, 2017)。

以物体为中心的场景加工理论过分强调了物体的作用, 而空间包裹理论过分看重全局特性的作用, 二者皆忽略了场景、物体、物体情境关系之间天然存在不可分割的空间和语义联系, 且缺乏相关神经机制的研究。随着研究的深入, 研究者逐步将 ERP 和 fMRI 技术应用到对场景加工机制的研究上来, 提出了场景联结加工假设。

4.3　场景的联结加工假设

场景联结加工(associative processing)假设有两点核心内容。首先, 大脑中储存了在长期经验中同时出现并已形成联结的重要场景信息, 而对场景的加工则是激活这种联结信息并加以提取的过程; 其次, 大脑的视觉系统中并不存在独立的场景加工区域, 这些已知场景选择性区域的作用实质是完成对外界信息中联结的加工, 这种联结信息可以是场景的整体特性、场景中物体间的情境关系等视觉信息, 也可以是听觉信息或其他感觉通道的信息(Aminoff & Tarr, 2015; Bar, Aminoff, Mason, & Fenske, 2007)。例如, 桌子和椅子在场景中经常一同出现, 大脑皮层可以根据经验将这两种刺激形成语义联结并储存在长时记忆中, 当新的外界刺激中出现了这种联结关系, 便会对长时记忆的语义联结信息进行激活, 利用这种联结信息完成对当前场景的加工。值得注意的是, 联结的形式并不是固定的, 可以是语义联结、空间关系联结或情绪联结等, 是在长期经验中通过学习而形成并稳定下来的(Bar et al., 2007)。

Aminoff 和 Tarr (2015)设计了一些由无意义图形组成的人工联结刺激: 保留图形间形状联合信息(图形类型不变, 位置变化, 黑色背景)的刺激集或者保留图形间空间联合信息(图形类型变化, 位置不变, 黑色背景)的刺激集, 通过学习让大脑对图形中两种联结信息纳入记忆, 在正式实验过程中要求被试判断刺激中包含了哪一种联结信息, 结果大脑皮层的场景选择区域对人工联结刺激的激活与真实场景类似, 表明大脑对人工联结刺激的加工与真实场景相似, 支持了场景选择区域的作用是对场景中的联结信息进行加工的观点。

场景联结加工假设为大量研究的结果提供了可能的解释。首先, 场景联结加工理论为物体与场景在加工过程中的相互促进提供了解释。一方面, 场景对处于其中物体的识别、搜索和记忆存在促进作用。例如, 处于一致物体间情境关系中的物体更容易被搜索和识别(如, Castelhano & Heaven, 2011; Davenport & Potter, 2004; 白学军, 康廷虎, 闫国利, 2008; 田宏杰, 王福兴, 徐菲菲, 申继亮, 2010); 对场景背景中的目标物体完成搜索任务和有意记忆任务, 搜索任务中目标物体的回忆效果更佳(Draschkow, Wolfe, & Võ, 2014; Josephs, Draschkow, Wolfe, & Võ, 2016)等; 另一方面, 已有研究发现物体也能促进对所处场景信息的加工。Davenport 和 Potter (2004)发现当物体与场景情境一致时, 被试对该场景识别的正确率提升。Wu, Wang 和 Pomplun (2014)将场景图片中的中心物体取出与灰色背景融合并打乱其空间关系, 要求被试完成对合成场景的分类任务, 结果发现任务的正确率显著高于平均水平。这表明在只保留物体间情境关系的情况下, 视觉系统仍然能够根据物体间的联结信息来完成场景的类别判断任务, 近期确有研究暗示场景的情境一致性确实会影响场景的分类(Collet, Fize, & VanRullen, 2015)。另外在 fMRI 研究中, 情境一致的场景和物体对 PPA 区域的激活更强(Bar & Aminoff, 2003; Troiani, Stigliani, Smith, & Epstein, 2014)。同时, Stansbury, Naselaris 和 Gallant (2013)发现视觉系统对场景分类的过程中物体出现的可能性参与了场景表征的形成。这些结果暗示, 场景与物体的加工并不是分离的, 而是相互关联共同完成的, 这种联合信息的存在为联结加工理论提供了支持。

其次, 场景联结加工理论为语义和句法不一致的 ERPs 研究结果提供了解释。Biederman,

Mezzanotte 和 Rabinowitz (1982)首次提出用语义和句法来描述不同的场景-物体的一致性关系，比如消防栓树立在街道旁属于语义一致，而悬浮在街道旁就属于语法不一致。将肥皂盒放在桌面上属于语法一致，但与笔记本电脑摆在一起就属于语义不一致。一项 ERPs 的研究结果发现场景与物体的语义不一致引发 N400 波，句法结构不一致引发 P600 波(Võ & Wolfe, 2013)，而对应的脑电成分在句子理解的语义与句法不一致情况下也会出现。如果假设大脑皮层对刺激的加工是对不同刺激之间联结的提取和激活，那么这种物体与背景不同联结方式(语义不一致/句法不一致)引起不同加工方式和脑电成分的现象就可以得到解释。N400 效应在多种形式的刺激中被发现，例如：语言、图片、物体、动作、声音等(Kutas & Federmeier, 2011)，这意味着不同类型刺激的语义加工存在一个共同的脑机制，那么语义/句法这种物体与场景信息联结方式的一致性也可能是不同刺激类型出现相同脑电成分的原因。

再次，场景联结加工理论为现有的场景选择区域对非场景刺激的强激活提供了解释。大脑皮层目前已知的场景选择区域是根据其对场景刺激区别于其他类型刺激所产生的反应偏向性确定而来的。然而随着研究的深入，一些场景选择性区域被发现参与了“非场景”类刺激的编码。以旁海马空间加工区(Parahippocampal Place Area, PPA)为例：PPA 区域对熟悉的面孔比不熟悉面孔的反应更大(Bar, Aminoff, & Ishai, 2008)，对具有高空间情境性的物体(如，车灯)比具有低空间情境性的物体(如，水壶)的激活更大(Bar, Aminoff, & Schacter, 2008)，对脱离背景的物体较面孔刺激的激活更大(Epstein & Kanwisher, 1998)等。场景选择性脑区对非场景类型刺激的反应体现了联结加工理论中“无独立场景加工脑区”的观点，即大脑对联结信息的加工并不根据信息类型的不同而做出区分。

5 场景主旨加工的神经基础

场景主旨加工的神经基础研究主要包括两个方面的内容：场景主旨加工的时间动力特性(时程特点)和场景主旨加工特异性的脑区。

5.1 时间动力特性

人类视觉系统能够迅速提取场景主旨的信息，对场景主旨加工的时间时程的研究往往采用较为短暂的呈现时间，要求记录生理信号的仪器设备需要有较高的时间分辨力，相关的研究工作主要是通过脑电和脑磁信号来完成的，目前这方面的研究相对较少。

Thorpe, Fize 和 Marlot (1996)率先采用脑电技术测量了视觉系统的加工速度，结果发现人类视觉系统在刺激呈现后的大约 150 ms 就可以将包含动物的场景与包含非动物的场景区分开来。随后 Sato 等人(1999)率先利用脑磁图(magnetoencephalography, MEG)技术，比较了大脑对场景和面孔的神经反应，以此探讨场景主旨加工的时间动力特性。结果发现，场景图片诱发的 MEG 信号的潜伏期(约 300 ms)长于面孔图片诱发的 MEG 信号的潜伏期(约 160 ms)。然而, Rivolta, Palermo, Schmalzl 和 Williams (2012)利用同样的技术，却发现场景刺激产生了与面孔同样早的类别特异性的 MEG 成分(M100p, 100~130 ms)。Bastin 等人(2013)采用脑电技术通过两个实验分离出神经活动中早期(200~500 ms)的刺激驱动效应和晚期(600~800 ms)的任务相关效应，并同时记录到了最强的 γ 波(50~150 Hz)。Groen 及其同事将场景图像的两个统计特征—空间相干(spatial coherence)和对比能量(energy contrast)与场景主旨加工联系起来，发现这两个统计指标均对早期(100~150 ms)单个试次的事件相关电位的波幅具有调节作用，且空间相干的影响可以持续到晚期(约 250 ms)的活动水平，对场景进行自然/人工的分类判断时，神经活动的幅度大小与空间相干有关，但与对比能量无关(Groen, Ghebreab, Prins, Lamme, & Scholte, 2013)。在后续的研究中，他们还发现虽然注意资源对场景刺激所诱发的早期(< 250 ms)电位活动几乎没有影响，但会对诱发的晚期(300~500 ms)电位活动产生较大的影响(Groen et al., 2016)。针对以往研究可能混淆了非场景刺激与场景刺激之间的物理差异所导致的神经信号的不同，最近的一项研究确定 P2 成分(潜伏期为约为 220 ms)是场景主旨加工的标记(Harel et al., 2016)。

5.2 特异脑区

相比较而言，对场景主旨加工特异性脑区的研究则探讨的是在更长时间尺度(数秒钟)上呈现场景刺激时，大脑不同区域的选择性激活，这类

研究要求仪器设备需要有较高的空间分辨力, 相关工作主要是通过 fMRI 技术来完成的, 这方面的研究相对较为丰富。

研究主要采用 fMRI 技术, 涉及到 3 个对场景类刺激具有反应偏向性的脑区: 海马旁区(parahippocampal place area, PPA) (如, Epstein & Kanwisher, 1998; Epstein, 2005)、压后皮层(retrosplenial complex, RSC) (Maguire, 2001)、枕区(occipital place area, OPA) (Dilks, Julian, Paunov, & Kanwisher, 2013), 以及一个对物体类刺激具有反应偏向的脑区: 枕叶外侧复合体(lateral occipital complex, LOC) (Malach et al., 1995)。

早期研究发现 PPA 与场景几何构型或空间布局(Rajimehr, Devaney, Bilenko, Young, & Tootell, 2011)以及空间延伸性(Kravitz, Peng, & Baker, 2011; Park, Brady, Greene, & Oliva, 2011)等特性的编码有关。这些研究认为 PPA 就像大脑中的“空间布局分析器”, 承担着场景空间信息表征建构的工作。后续研究发现 PPA 区域还与场景中物体信息的加工有关, 包括物体大小(如, Cant & Xu, 2012; Konkle & Oliva, 2012)、物体引发场景联想的难易(Mullally & Maguire, 2011)和场景分类(Dilks, Julian, Kubilius, Spelke, & Kanwisher, 2011; Peelen, Fei-Fei, & Kastner, 2009; Walther, Caddigan, Fei-Fei, & Beck, 2009)等。这些发现表明 PPA 不仅是一个“空间布局分析器”, 而且对场景中的物体也具有强的敏感性。

压后皮层(retrosplenial cortex, RSC)也是具有场景反应偏向性的重要脑区, 并与视觉系统的空间定位和导航能力有关(Epstein, Parker, & Feiler, 2007; Marchette, Vass, Ryan, & Epstein, 2014; Vann, Aggleton, & Maguire, 2009; Vass & Epstein, 2013)。RSC 与 PPA 且都与场景的空间延伸性有关(Henderson, Zhu, & Larson, 2011), 但 PPA 能够分辨出场景中细节的改变, 却不能判断视角改变前后的两个场景是否是同一个位置, 而 RSC 则能够将不同视角的变化与不同场景的变化区别开。PPA 与 RSC 的这种功能上的互补帮助视觉系统建立起完整且细节丰富的多方位场景表征(Park & Chun, 2009)。

不同于 PPA 和 RSC, 人们对 OPA 区域在场景加工中的功能知之甚少。Dilks 等人(2013)认为 OPA 可能是最先获得感受器传入的场景信息的脑区, 而 Silson, Chan, Reynolds, Kravitz 和 Baker (2015)认为 PPA 与 OPA 分别具有的上下视野偏差(upper and lower visual field biases)可能是两者在功能上存在互补性的证据。除此之外, OPA 与其他两个场景选择性脑区在场景加工中的反应偏向性有许多相似之处, 例如, 三者都具有熟悉性偏向(Epstein, Higgins, Jablonski, & Feiler, 2007)、直线偏向(Nasr, Echavarria, & Tootell, 2014)等。目前关于 OPA 脑区的更多功能仍然在探索之中, 关于 3 个场景选择脑区的反应偏向性和 OPA 的真正功能还需要进一步的实验验证。

枕叶外侧复合体(lateral occipital complex, LOC)最初由于其对物体的反应偏向性而被发现, 并被认为与物体形状和类别的编码有关(如, Eger, Ashburner, Haynes, Dolan, & Rees, 2008), 近来人们发现其在场景加工过程中也承担着重要作用。Walther 等人(2009)发现在此过程中 PPA、RSC 和 LOC 在协同作用下完成了对场景信息的加工。Harel, Kravitz 和 Baker (2013)发现, RSC 和 PPA 同时对场景的空间布局敏感, 而 LOC 和 PPA 则同时对场景中的物体敏感。

那么, 这些脑区究竟是怎样在场景加工的过程中协同完成任务的呢? PPA 在空间信息和物体属性加工上的双重作用应如何解释呢? 为了探索这种协同运作的具体方式, Baldassano, Beck 和 Fei-Fei (2013)采用功能联结分析(functional connectivity analysis)技术, 发现 PPA 区域的前部与 RSC 区域反应相关性很大, 而 PPA 区域的后部则与 LOC 区域反应相关性很大, 证明 PPA 是由前后两个功能不一致的部分组成: 前部与场景的记忆和情境加工有关, 而后部与低水平特征和物体形状的加工有关。近期一项研究发现, 视觉系统在针对不同空间布局、纹理材料, 以及具有开放或封闭、自然或人工属性的材料完成不同种类的判断任务时, PPA、RSC 和 OPA 区域的激活程度受到任务种类和场景属性交互作用的影响, 证明视觉系统虽然会对场景的多种信息进行提取和整合, 但这个过程会随着任务目的和场景特征的不同而变化(Lowe, Gallivan, Ferber, & Cant, 2016)。这样的结果与联结加工理论对场景加工的解释一致, 为未来对场景加工机制的探索提供了可行方向。

综上可知, PPA 的前部与 RSC 主要参与场景

的空间信息编码, PPA 后部与 LOC 主要参与场景中物体的加工, 而 OPA 的真正功能以及这些脑区在场景加工中的协同工作方式仍不清楚。目前有研究认为主旨表征是场景空间信息编码不可缺少的信息(Siddiqui & Brown, 2015), 其中场景的快速分类任务也与这些脑区相关(Ramkumar, Hansen, Pannasch, & Loschky, 2016), 但究竟它们在场景主旨加工中起到哪些具体作用? 这些问题的答案仍需要进一步的探索研究。

6 小结与展望

本文对真实场景主旨加工的影响因素、相关的争议性问题、加工机制的理论解释以及场景主旨加工的神经基础等方面进行评述。尽管该领域的相关研究已经取得了一些重要性的成果, 但这些研究成果中仍然存在一些亟待解决重要问题。

首先, 目前的研究结果支持场景主旨的识别需要全局特性作为诊断性信息, 而颜色、空间频率、线条朝向等低水平信息对主旨的加工也具有一定影响。据此可以对场景主旨的加工进行两个方面的研究：(1)已有研究表明视觉系统对真实场景的信息加工具有一定程度的灵活性, 这种灵活性在多大程度上影响了场景早期加工的信息提取仍不清楚, 因此可通过在同一实验范式中操纵任务或者经验等变量来研究它们对信息提取的影响, 以此来探讨视觉信息的前馈(feedforward)和反馈(feedback)加工之间的相互关系(Maguire & Howe, 2016; Potter et al., 2014; 孙雨生等, 2017); (2)已有研究对全局特征的种类表述不够规范, 因此可以进一步探讨是否存在场景信息加工的基本单元, 究竟哪些特征(如自然性、开放性、粗糙度等)可以作为场景信息加工的基本单元, 这些特征对于不同类型的场景图像是否具有特异性, 此方向的研究有利于找到更好的计算机分类算法, 使机器视觉更加接近人类的视觉系统。(3)场景图像中全局特征的加工与群集表征(assembly representation)加工之间有什么样的关系, 二者的加工是否具有共同的机制(Alvarez, 2011; Brady, Shafer-Skelton, & Alvarez, 2017; Cohen, Dennett, & Kanwisher, 2016; De Cesarei, Loftus, Mastria, & Codispoti, 2017)。

其次, 就场景主旨加工的理论解释而言, 我们认为一般情况下主旨是以场景的全局特征作为诊断性信息的, 而场景主旨加工会影响注意引导和记忆(Malcolm et al., 2016)。但场景中的物体是否会影响, 如何影响场景主旨的加工仍不清楚。为了回答该问题, 未来研究可以从以下两个方面展开研究：(1)已有研究表明场景与物体的语义一致性对物体的加工具有促进作用, 这种一致性是否能够促进场景主旨的加工以及这种一致性判断在场景主旨识别过程中是否为必要过程; (2)已有研究表明场景主旨与物体加工存在交互作用, 这种交互作用是否暗示场景与物体加工并不是平行过程而是相互促进、相互制约, 如果是, 那么这种交互作用是怎样在注意的调节下完成的。

第三, 目前的研究结果在场景的层级加工优先性上存在矛盾: 早期研究者发现了场景分类基本水平优势; 然而后来的研究采用不同的实验范式和方法获得了上级水平优势。我们认为这种看似存在矛盾的层级加工优先性的模式可能并不那么稳定, 会受到一些因素的调节, 未来研究可从以下几个方面做进一步的探讨: (1)近期已有研究探讨了场景图片的相似性对层级加工优先性的影响(Banno & Saiki, 2015, Poncet & Fabre-Thorpe, 2014); 后续研究可以探讨其他调节这种层级加工优先性的边界条件, 如场景图片的典型性等。(2)已有研究表明工作记忆内容能够增强视知觉表征(如, Soto, Wriglesworth, Bahrami-Balani, & Humphreys, 2010); 后续研究可以探讨工作记忆的内容对场景层级加工优先性的调节作用。(3)有研究表明奖赏能够改变知觉(Failing & Theeuwes, 2016), 那么场景中包含的具有生存价值/情绪性意义的信息是否也会影响场景主旨加工的层级优先性也是一个值得探讨的问题(李毕琴等, 2015)。另外, 还有研究暗示场景主旨的加工具有个体差异性, 那么个体差异是否也会影响场景主旨层级加工优先性值得进一步的研究。对这类问题的研究将有助于揭示场景主旨加工的时间进程。

第四, 目前的研究对场景主旨的加工或分类是否需要注意资源的问题尚未有定论。当前对场景主旨是否需要注意参与的问题可以从以下方向进行研究：(1) Groen 等人(2016)发现除了任务难度, 在双任务条件下的两种目标刺激在空间上是否发生重叠可能会影响到两种任务在注意分配上是否存在竞争, 这一假设有待进一步考证; (2) Kay, Weiner 和 Grill-Spector (2015)在一项 fMRI

研究中使用人脸作为刺激材料发现注意状态对早期视觉皮层的激活没有影响，仅仅影响晚期高水平类别选择区域(high-level category selective area)，表明人脸刺激中的早期低水平信息加工不易受注意资源多少的影响。在这一研究的启示下，可分开探讨场景不同水平信息加工与注意资源的关系，即是否场景的部分信息需要注意参与才能完成提取和加工，而另一些信息的加工则不需要注意参与。

第五，关于场景主旨加工的时间动力特性的神经基础方面，目前的研究还相对较为薄弱，结果也不尽相同，还需要设计巧妙的实验，利用高时间分辨率的 ERP 技术来明确场景主旨加工的时间进程。在场景主旨加工的特异性脑区方面的研究较为丰富，大多数研究主要提及了 PPA、RSC、OPA 和 LOC 四个区域在场景主旨加工中的作用，但对这些区域在场景主旨加工过程中是如何协同作用的研究相对较少，未来研究可以考虑利用脑功能联结分析技术探讨这些区域的协同作用，为场景知觉加工建立相应的脑功能网络。这个领域的研究对于智能机器视觉的研制和深度神经网络的构建具有启示意义。

到目前为止，场景主旨加工的研究虽然取得了丰硕的成果，但还存在诸多争议性的问题亟待解决。我们期待将来有更多研究者加入到该领域的研究中来，将行为实验、眼动技术、计算建模和现代化脑成像技术等多种方法相融合，对场景主旨加工进行全面立体的研究，以丰富人们对真实场景加工过程的理解。

参考文献

白学军，康廷虎，闫国利. (2008). 真实情景中刺激物识别的理论模型与研究回顾. *心理科学进展, 16*(5), 679–686.

程昊. (2010). *复杂场景的主旨提取机制* (硕士学位论文). 浙江大学，杭州.

李毕琴，郭毕鹏，胡竹菁，罗跃嘉. (2015). 场景主旨对恐惧刺激加工影响的 ERP 研究. *心理学探新, 35*(4), 318–323.

孙琪，任衍具，傅根跃. (2015). 视空工作记忆负载对场景主旨加工的影响. *心理科学, 38*(6), 1311–1318.

孙雨生，张智君，吴彬星. (2017). 上下文预期在快速场景识别中的作用. *心理学报, 49*(5), 577–589.

田宏杰，王福兴，徐菲菲，申继亮. (2010). 场景知觉中物体加工的背景效应. *心理科学进展, 18*(6), 878–886.

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences, 15*(3), 122–131.

Aminoff, E. M., & Tarr, M. J. (2015). Associative processing is inherent in scene perception. *PLoS One, 10*(6), e0128840.

Awasthi, B., Sowman, P. F., Friedman, J., & Williams, M. A. (2013). Distinct spatial scale sensitivities for early categorization of faces and places: Neuromagnetic and behavioral findings. *Frontiers in Human Neuroscience, 7*, 91.

Baldassano, C., Beck, D. M., & Fei-Fei, L. (2013). Differential connectivity within the parahippocampal place area. *NeuroImage, 75*, 228–237.

Banno, H., & Saiki, J. (2015). The processing speed of scene categorization at multiple levels of description: The superordinate advantage revisited. *Perception, 44*(3), 269–288.

Bar, M., & Aminoff, E. (2003). Cortical analysis of visual context. *Neuron, 38*(2), 347–358.

Bar, M., Aminoff, E., Mason, M., & Fenske, M. (2007). The units of thought. *Hippocampus, 17*, 420–428.

Bar, M., Aminoff, E., & Ishai, A. (2008). Famous faces activate contextual associations in the parahippocampal cortex. *Cerebral Cortex, 18*(6), 1233–1238.

Bar M., Aminoff, E., & Schacter, D. L. (2008). Scenes unseen: The parahippocampal cortex intrinsically subserves contextual associations, not scenes or places per se. *Journal of Neuroscience, 28*(34), 8539–8544.

Bastin, J., Committeri, G., Kahane, P., Galati, G., Minotti, L., Lachaux, J. P., & Berthoz, A. (2013). Timing of posterior parahippocampal gyrus activity reveals multiple scene processing stages. *Human Brain Mapping, 34*(6), 1357–1370.

Biederman, I. (1972). Perceiving real-world scenes. *Science, 177*(4043), 77–80.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology, 14*(2), 143–177.

Biggs, A. T., & Mitroff, S. R. (2015). Improving the efficacy of security screening tasks: A review of visual search challenges and ways to mitigate their adverse effects. *Applied Cognitive Psychology, 29*(1), 142–148.

Boucart, M., Moroni, C., Thibaut, M., Szaffarczyk, S., & Greene, M. (2013). Scene categorization at large visual eccentricities. *Vision Research, 86*, 35–42.

Brady, T. F., & Oliva, A. (2008). Statistical learning using real-world scenes: Extracting categorical regularities without conscious intent. *Psychological Science, 19*(7), 678–685.

Brady, T. F., Shafer-skelton, A., & Alvarez, G. A. (2017).

Global ensemble texture representations are critical to rapid scene perception. *Journal of Experimental Psychology Human Perception & Performance, 43*(6), 1160–1176.

Brand, J., & Johnson, A. P. (2014). Attention to local and global levels of hierarchical Navon figures affects rapid scene categorization. *Frontiers in Psychology, 5*, 1274.

Cant, J. S., & Xu, Y. (2012). Object ensemble processing in human anterior-medial ventral visual cortex. *Journal of Neuroscience, 32*(22), 7685–7700.

Castelhano, M. S., & Henderson, J. M. (2008). The influence of color on the perception of scene gist. *Journal of Experimental Psychology: Human Perception and Performance, 34*(3), 660–675.

Castelhano, M. S., & Heaven, C. (2011). Scene context influences without scene gist: Eye movements guided by spatial associations in visual search. *Psychonomic Bulletin & Review, 18*(5), 890–896.

Clarke, J., & Mack, A. (2014). Iconic memory for the gist of natural scenes. *Consciousness and Cognition, 30*, 256–265.

Cohen, M. A., Alvarez, G. A., & Nakayama, K. (2011). Natural-scene perception requires attention. *Psychological Science, 22*(9), 1165–1172.

Cohen, M. A., Dennett, D. C., & Kanwisher, N. (2016). What is the bandwidth of perceptual experience? *Trends in Cognitive Sciences, 20*(5), 324–335.

Collet, A.C., Fize, D., & VanRullen, R. (2015). Contextual congruency effect in natural scene categorization: Different strategies in humans and monkeys (*Macaca mulatta*). *PLoS One, 10*(7), e0133721.

Csathó, Á., van der Linden, D., & Gács, B. (2015). Natural scene recognition with increasing time-on-task: The role of typicality and global image properties. *The Quarterly Journal of Experimental Psychology, 68*(4), 814–828.

Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science, 15*(8), 559–564.

De Graef, P., Christiaens, D., & D'Ydewalle, G. (1990). Perceptual effects of scene context on object identification. *Psychological Research, 52*(4), 317–329.

De Cesarei, A., Loftus, G. R., Mastria, S., & Codispoti, M. (2017). Understanding natural scenes: Contributions of image statistics. *Neuroscience & Biobehavioral Reviews, 74*(Part A), 44–57.

Dilks, D. D., Julian, J. B., Kubilius, J., Spelke, E. S., & Kanwisher, N. (2011). Mirror-image sensitivity and invariance in object and scene processing pathways. *Journal of Neuroscience, 31*(31), 11305–11312.

Dilks, D. D., Julian, J. B., Paunov, A. M., & Kanwisher, N. (2013). The occipital place area is causally and selectively involved in scene perception. *Journal of Neuroscience, 33*(4), 1331–1336.

Draschkow, D., Wolfe, J. M., & Võ, M. L.-H. (2014). Seek and you shall remember: Scene semantics interact with visual search to build better memories. *Journal of Vision, 14*(8), 10.

Duh, S., & Wang, S.-H. (2014). Infants detect changes in everyday scenes: The role of scene gist. *Cognitive Psychology, 72*, 142–161.

Eger, E., Ashburner, J., Haynes, J. D., Dolan, R. J., & Rees, G. (2008). fMRI activity patterns in human LOC carry information about object exemplars within category. *Journal of Cognitive Neuroscience, 20*(2), 356–370.

Epstein, R. (2005). The cortical basis of visual scene processing. *Visual Cognition, 12*(6), 954–978.

Epstein, R. A., Higgins, J. S., Jablonski, K., & Feiler, A. M. (2007). Visual scene processing in familiar and unfamiliar environments. *Journal of Neurophysiology, 97*(5), 3670–3683.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature, 392*(6676), 598–601.

Epstein, R. A., Parker, W. E., & Feiler, A. M. (2007). Where am I now? Distinct roles for parahippocampal and retrosplenial cortices in place recognition. *Journal of Neuroscience, 27*(23), 6141–6149.

Evans, K. K., Haygood, T. M., Cooper, J., Culpan, A.-M., & Wolfe, J. M. (2016). A half-second glimpse often lets radiologists identify breast cancer cases even when viewing the mammogram of the opposite breast. *Proceedings of the National Academy of Sciences of the United States of America, 113*(37), 10292–10297.

Fabre-Thorpe, M. (2011). The characteristics and limits of rapid visual categorization. *Frontiers in Psychology, 2*, 243.

Failing, M., & Theeuwes, J. (2016). Reward alters the perception of time. *Cognition, 148*, 19–26.

Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). What do we perceive in a glance of a real-world scene? *Journal of Vision, 7*(1), 10.

Freeman, T. E., Loschky, L. C., & Hansen, B. C. (2015). Scene masking is affected by trial blank-screen luminance. *Signal Processing: Image Communication, 39*(Part B), 319–327.

Friedman, A. (1979). Framing pictures: The role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology: General, 108*(3), 316–355.

Fu, Q. F., Liu, Y.-J., Dienes, Z., Wu, J. H., Chen, W. F., & Fu, X. L. (2016). The role of edge-based and surface-based information in natural scene categorization: Evidence from behavior and event-related potentials. *Consciousness*

and Cognition, 43, 152–166.

Gegenfurtner, K. R., & Rieger, J. (2000). Sensory and cognitive contributions of color to the recognition of natural scenes. *Current Biology, 10*(13), 805–808.

Goffaux, V., Jacques, C., Mouraux, A., Oliva, A., Schyns, P. G., & Rossion, B. (2005). Diagnostic colours contribute to the early stages of scene categorization: Behavioural and neurophysiological evidence. *Visual Cognition, 12*(6), 878–892.

Goldzieher, M. J., Andrews, S., & Harris, I. M. (2017). Two scenes or not two scenes: The effects of stimulus repetition and view-similarity on scene categorization from brief displays. *Memory & Cognition, 45*(1), 49–62.

Greene, M. R., & Oliva, A. (2006, July). *Natural scene categorization from conjunctions of ecological global properties*. Proceedings of the 28th Annual Conference of the Cognitive Science Society (pp. 291–296), Vancouver, Canada.

Greene, M. R., & Oliva, A. (2009a). Recognition of natural scenes from global properties: Seeing the forest without representing the trees. *Cognitive Psychology, 58*(2), 137–176.

Greene, M. R., & Oliva, A. (2009b). The briefest of glances: The time course of natural scene understanding. *Psychological Science, 20*(4), 464–472.

Greene, M. R., & Fei-Fei, L. (2014). Visual categorization is automatic and obligatory: Evidence from Stroop-like paradigm. *Journal of Vision, 14*(1), 14.

Greene, M. R., Botros, A. P., Beck, D. M., & Fei-Fei, L. (2015). What you see is what you expect: Rapid scene understanding benefits from prior experience. *Attention, Perception & Psychophysics, 77*(4), 1239–1251.

Groen, I. I. A., Ghebreab, S., Lamme, V. A. F., & Scholte, H. S. (2016). The time course of natural scene perception with reduced attention. *Journal of Neurophysiology, 115*(2), 931–946.

Groen, I. I. A., Ghebreab, S., Prins, H., Lamme, V. A. F., & Scholte, H. S. (2013). From image statistics to scene gist: Evoked neural activity reveals transition from low-level natural image structure to scene category. *The Journal of Neuroscience, 33*(48), 18814–18824.

Gronau, N., & Izoutcheev, A. (2017). The necessary of visual attention to scene categorization: Dissociating ‘task-relevant’ and ‘task-irrelevant’ scene distractors. *Journal of Experimental Psychology: Human Perception and Performance, 43*(5), 954–970.

Hagmann, C. E., & Potter, M. C. (2016). Ultrafast scene detection and recognition with limited visual information. *Visual Cognition*, *24*(1), 2–14.

Hansen, B. C., & Loschky, L. C. (2013). The contribution of amplitude and phase spectra-defined scene statistics to the masking of rapid scene categorization. *Journal of Vision, 13*(13), 21.

Harel, A., Kravitz, D. J., & Baker, C. I. (2013). Deconstructing visual scenes in cortex: Gradients of object and spatial layout information. *Cerebral Cortex, 23*(4), 947–957.

Harel, A., Groen, I. I. A., Kravitz, D. J., Deouell, L. Y., & Baker, C. I. (2016). The temporal dynamics of scene processing: A multifaceted EEG investigation. *eNeuro, 3*(5), e0139-16.

Henderson, J. M., Zhu, D. C., & Larson, C. L. (2011). Functions of parahippocampal place area and retrosplenial cortex in real-world scene analysis: An fMRI study. *Visual Cognition, 19*(7), 910–927.

Josephs, E. L., Draschkow, D., Wolfe, J. M., & Võ, M. L.-H. (2016). Gist in time: Scene semantics and structure enhance recall of searched objects. *Acta Psychologica, 169*, 100–108.

Joubert, O. R., Rousselet, G. A., Fize, D., & Fabre-Thorpe, M. (2007). Processing scene context: Fast categorization and object interference. *Vision Research, 47*(26), 3286–3297.

Joubert, O. R., Rousselet, G. A., Fabre-Thorpe, M., & Fize, D. (2009). Rapid visual categorization of natural scene contexts with equalized amplitude spectrum and increasing phase noise. *Journal of Vision, 9*(1), 2.

Kadar, I., & Ben-Shahar, O. (2012). A perceptual paradigm and psychophysical evidence for hierarchy in scene gist processing. *Journal of Vision, 12*(13), 16.

Kauffmann, L., Chauvin, A., Pichat, C., & Peyrin, C. (2015). Effective connectivity in the neural network underlying coarse-to-fine categorization of visual scenes. A dynamic causal modeling study. *Brain & Cognition, 99*, 46–56.

Kauffmann, L., Ramanoël, S., & Peyrin, C. (2014). The neural bases of spatial frequency processing during scene perception. *Frontiers in Integrative Neuroscience, 8*, 37.

Kay, K. N., Weiner, K. S., & Grill-Spector, K. (2015). Attention reduces spatial uncertainty in human ventral temporal cortex. *Current Biology, 25*(5), 595–600.

Kihara, K., & Takeda, Y. (2010). Time course of the integration of spatial frequency-based information in natural scenes. *Vision Research, 50*(21), 2158–2162.

Kihara, K., & Takeda, Y. (2012). Attention-free integration of spatial frequency-based information in natural scenes. *Vision Research, 65*, 38–44.

Konkle, T., & Oliva, A., (2012). A real-world size organization of object responses in occipitotemporal cortex. *Neuron, 74*(6), 1114–1124.

Kravitz, D. J., Peng, C. S., & Baker, C. I. (2011). Real-world

scene representations in high-level visual cortex: It's the spaces more than the places. *The Journal of Neuroscience, 31*(20), 7322–7333.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology, 62*(1), 621–647.

Kveraga, K., Boshyan, J., & Bar, M. (2007). Magnocellular projections as the trigger of top-down facilitation in recognition. *Journal of Neuroscience, 27*(48), 13232–13240.

Larson, A. M., & Loschky, L. C. (2009). The contributions of central versus peripheral vision to scene gist recognition. *Journal of Vision, 9*(10), 6.

Larson, A. M., Freeman, T. E., Ringer, R. V., & Loschky, L. C. (2014). The spatiotemporal dynamics of scene gist recognition. *Journal of Experimental Psychology: Human Perception and Performance, 40*(2), 471–487.

Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences of the United States of America, 99*(14), 9596–9601.

Loschky, L. C., Hansen, B. C., Sethi, A., & Pydimarri, T. N. (2010). The role of higher order image statistics in masking scene gist recognition. *Attention, Perception, & Psychophysics, 72*(2), 427–444.

Loschky, L. C., & Larson, A. M. (2010). The natural/man-made distinction is made before basic-level distinctions in scene gist processing. *Visual Cognition, 18*(4), 513–536.

Loschky, L. C., Ringer, R. V., Ellis, K., & Hansen, B. (2015). Comparing rapid scene categorization of aerial and terrestrial views: A new perspective on scene gist. *Journal of Vision, 15*(6), 11.

Loschky, L. C., Nuthmann, A., Fortenbaugh, F. C., & Levi, D. M. (2017). Scene perception from central to peripheral vision. *Journal of Vision, 17*(1), 6.

Lowe, M. X., Gallivan, J. P., Ferber, S., & Cant, J. S. (2016). Feature diagnosticity and task context shape activity in human scene-selective cortex. *NeuroImage, 125*, 681–692.

Macé, M. J.-M., Joubert, O. R., Nespoulous, J. L., & Fabre-Thorpe, M. (2009). The time-course of visual categorizations: You spot the animal faster than the bird. *PLoS One, 4*(6), e5927.

Mack, A., & Clarke, J. (2012). Gist perception requires attention. *Visual Cognition, 20*(3), 300–327.

Mack, M. L., & Palmeri, T. J. (2015). The dynamics of categorization: Unraveling rapid categorization. *Journal of Experimental Psychology: General, 144*(3), 551–569.

Maguire, E. A. (2001). The retrosplenial contribution to human navigation: A review of lesion and neuroimaging findings. *Scandinavian Journal of Psychology, 42*(3), 225–238.

Maguire, J. F., & Howe, P. D. L. (2016). Failure to detect meaning in RSVP at 27 ms per picture. *Attention, Perception, & Psychophysics, 78*(5), 1405–1413.

Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H, Kennedy, W. A., … Tootell, R. B. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Science of the United States of America, 92*(18), 8135–8139.

Malcolm, G. L., Nuthmann, A., & Schyns, P. G. (2014). Beyond gist: Strategic and incremental information accumulation for scene categorization. *Psychological Science, 25*(5), 1087–1097.

Malcolm, G. L., Groen, I. I. A., & Baker, C. I. (2016). Making sense of real-world scenes. *Trends in Cognitive Sciences, 20*(11), 843–856.

Marchette, S. A., Vass, L. K., Ryan, J., & Epstein, R. A. (2014). Anchoring the neural compass: Coding of local spatial reference frames in human medial parietal lobe. *Nature Neuroscience, 17*(11), 1598–1606.

Marx, S., Hansen-Goos, O., Thrun, M. C., & Einhäuser, W. (2014). Rapid serial processing of natural scenes: Color modulates detection but neither recognition nor the attentional blink. *Journal of Vision, 14*(14), 4.

Morrison, D. J., & Schyns, P. G. (2001). Usage of spatial scales for the categorization of faces, objects, and scenes. *Psychonomic Bulletin & Review, 8*(3), 454–469.

Mu, T. L., & Li, S. (2013). The neural signature of spatial frequency-based information integration in scene perception. *Experimental Brain Research, 227*(3), 367–377.

Mullally, S. L., & Maguire, E. A. (2011). A new role for the parahippocampal cortex in representing space. *Journal of Neuroscience, 31*(20), 7441–7449.

Musel, B., Chauvin, A., Guyader, N., Chokron, S., & Peyrin, C. (2012). Is coarse-to-fine strategy sensitive to normal aging? *PLos One, 7*(6), e38493.

Musel, B., Kauffmann L., Ramanoël, S., Giavarini, C., Guyader, N., Chauvin, A., & Peyrin, C. (2014). Coarse-to-fine categorization of visual scenes in scene-selective cortex. *Journal of Cognitive Neuroscience, 26*(10), 2287–2297.

Nasr, S., Echavarria, C. E., & Tootell, R. B. H. (2014). Thinking outside the box: Rectilinear shapes selectively activate scene-selective cortex. *Journal of Neuroscience, 34*(20), 6721–6735.

Oliva, A., & Schyns, P. G. (1997). Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology, 34*(1), 72–107.

Oliva, A. (2005). Gist of the scene. In L. Itti, G. Rees, & J. K. Tsotsos (Eds.), *Neurobiology of attention* (pp. 251–256). San Diego, CA: Elsevier.

Oliva, A., & Schyns, P. G. (2000). Diagnostic colors mediate scene recognition. *Cognitive Psychology, 41*(2), 176–210.

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A global representation of the spatial envelope. *International Journal of Computer Vision, 42*(3), 145–175.

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research, 155*, 23–36.

Otsuka, S., & Kawaguchi, J. (2007). Natural scene categorization with minimal attention: Evidence from negative priming. *Perception & Psychophysics, 69*(7), 1126–1139.

Otsuka, S., & Kawaguchi, J. (2009). Direct versus indirect processing changes the influence of color in natural scene categorization. *Attention, Perception & Psychophysics, 71*(7), 1588–1597.

Otsuka, Y., Ichikawa, H., Kanazawa, S., Yamaguchi, M. K., & Spehar, B. (2014). Temporal dynamics of spatial frequency processing in infants. *Journal of Experimental Psychology: Human Perception and Performance, 40*(3), 995–1008.

Park, S., & Chun, M. M. (2009). Different roles of the parahippocampal place area (PPA) and retrosplenial cortex (RSC) in panoramic scene perception. *NeuroImage, 47*(4), 1747–1756.

Park, S., Brady, T. F., Greene, M. R., & Oliva, A. (2011). Disentangling scene content from spatial boundary: Complementary roles for the parahippocampal place area and lateral occipital complex in representing real-world scenes. *Journal of Neuroscience, 31*(4), 1333–1340.

Patai, E. Z., Buckley, A., & Nobre, A. C. (2013). Is attention based on spatial contextual memory preferentially guided by low spatial frequency signals? *PLos One, 8*(6), e65601.

Peelen, M. V., Fei-Fei, L., & Kastner, S. (2009). Neural mechanisms of rapid natural scene categorization in human visual cortex. *Nature, 460*(7251), 94–97.

Poncet, M., Reddy, L., & Fabre-Thorpe, M. (2012). A need for more information uptake but not focused attention to access basic-level representations. *Journal of Vision, 12*(1), 15.

Poncet, M., & Fabre-Thorpe, M. (2014). Stimulus duration and diversity do not reverse the advantage for superordinate-level representations: The animal is seen before the bird. *European Journal of Neuroscience, 39*(9), 1508–1516.

Potter, M. C. (1975). Meaning in visual search. *Science, 187*(4180), 965–966.

Potter, M. C., & Levy, E. I. (1969). Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology, 81*(1), 10–15.

Potter, M. C., Wyble, B., Hagmann, C. E., & McCourt, E. S. (2014). Detecting meaning in RSVP at 13 ms per picture. *Attention, Perception, & Psychophysics, 76*(2), 270–279.

Rajimehr, R., Devaney, K. J., Bilenko, N. Y., Young, J. C., & Tootell, R. B. H. (2011). The "parahippocampal place area" responds preferentially to high spatial frequencies in humans and monkeys. *PLoS Biology, 9*(4), e1000608.

Ramkumar, P., Hansen, B. C., Pannasch, S., & Loschky, L. C. (2016). Visual information representation and rapid-scene categorization are simultaneous across cortex: An MEG study. *NeuroImage, 134*, 295–304.

Rivolta, D., Palermo, R., Schmalzl, L., & Williams, M. A. (2012). An early category-specific neural response for the perception of both places and faces. *Cognitive Neuroscience, 3*(1), 45–51.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*(3), 382–439.

Rotshtein, P., Schofield, A., Funes, M. J., & Humphreys, G. W. (2010). Effects of spatial frequency bands on perceptual decision: It is not the stimuli but the comparison. *Journal of Vision, 10*(10), 25.

Rousselet, G., Joubert, O., & Fabre-Thorpe, M. (2005). How long to get the "gist" of real-world natural scene? *Visual Cognition, 12*(6), 852–877.

Sato, N., Nakamura, K., Nakamura, A., Sugiura, M., Ito, K., Fukuda, H., & Kawashima, R. (1999). Different time course between scene processing and face processing: An MEG study. *Neuroreport, 10*(17), 3633–3637.

Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time- and spatial-scale-dependent scene recognition. *Psychological Science, 5*(4), 195–200.

Serre, T. (2016). Models of visual categorization. *Wiley Interdisciplinary Reviews: Cognitive Science, 7*(3), 197–213.

Siddiqui, A., & Brown, J. (2015). The role of gist processing in boundary extension. *Journal of Vision, 15*(12), 354.

Silson, E. H., Chan, A. W.-Y., Reynolds, R. C., Kravitz, D. J., & Baker, C. I. (2015). A retinotopic basis for the division of high-level scene processing between lateral and ventral human occipitotemporal cortex. *Journal of Neuroscience, 35*(34), 11921–11935.

Sofer, I., Crouzet, S. M., & Serre, T. (2015). Explaining the

timing of natural scene understanding with a computational model of perceptual categorization. *PLoS Computational Biology, 11*(9), e1004456.

Soto, D., Wriglesworth, A., Bahrami-Balani, A., & Humphreys, G. W. (2010). Working memory enhances visual perception: Evidence from signal detection analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(2), 441–456.

Stansbury, D. E., Naselaris, T., & Gallant, J. L. (2013). Natural scene statistics account for the representation of scene categories in human visual cortex. *Neuron, 79*(5), 1025–1034.

Subramanian, R., Shankar, D., Sebe, N., & Melcher, D. (2014). Emotion modulates eye movement patterns and subsequent memory for the gist and details of movie scenes. *Journal of Vision, 14*(3), 31.

Sun, Q., Ren, Y. J., Zheng, Y., Sun, M. X., & Zheng, Y. J. (2016). Superordinate level processing has priority over basic-level processing in scene gist recognition. *i-Perception, 7*(6), 1–18.

Thibaut, M., Tran, T. H. C., Szaffarczyk, S., & Boucart, M. (2014). The contribution of central and peripheral vision in scene categorization: A study on people with central vision loss. *Vision Research, 98*, 46–53.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*(6582), 520–522.

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review, 113*(4), 766–786.

Troiani, V., Stigliani, A., Smith, M. E., & Epstein, R. A. (2014). Multiple object properties drive scene-selective regions. *Cerebral Cortex, 24*(4), 883–897.

Tversky, B., & Hemenway, K. (1983). Categories of environmental scenes. *Cognitive Psychology, 15*(1), 121–149.

Vanmarcke, S., & Wagemans, J. (2015). Rapid gist perception of meaningful real-life scenes: Exploring individual and gender differences in multiple categorization tasks. *i-Perception, 6*(1), 19–37.

Vanmarcke, S., & Wagemans, J. (2016). Individual differences in spatial frequency processing in scene perception: The influence of autism-related traits. *Visual Cognition, 24*(2), 115–131.

Vanmarcke, S., van Esch, L., van der Hallen, R., Evers, K., Noens, I., Steyaert, J., & Wagemans, J. (2016). Gist perception in adolescents with and without ASD: Ultra-rapid categorization of meaningful real-life scenes. *Research in Autism Spectrum Disorders, 29-30*, 30–47.

Vann, S. D., Aggleton, J. P., & Maguire, E. A. (2009). What does the retrosplenial cortex do? *Nature Reviews Neuroscience, 10*(11), 792–802.

Vass, L. K., & Epstein, R. A. (2013). Abstract representations of location and facing direction in the human brain. *Journal of Neuroscience, 33*(14), 6133– 6142.

Võ, M. L.-H., & Wolfe, J. M. (2013). Differential electrophysiological signatures of semantic and syntactic scene processing. *Psychological Science, 24*(9), 1816–1823.

Walker, S., Stafford, P., & Davis, G. (2008). Ultra-rapid categorization requires visual attention: Scenes with multiple foreground objects. *Journal of Vision, 8*(4), 21.

Walther, D. B., Caddigan, E., Fei-Fei, L., & Beck, D. M. (2009). Natural scene categories revealed in distributed patterns of activity in the human brain. *Journal of Neuroscience, 29*(34), 10573–10581.

Walther, D. B., & Shen, D. D. (2014). Nonaccidental properties underlie human categorization of complex natural scenes. *Psychological Science, 25*(4), 851–860.

Wang, P. Q., & Cottrell, G. (2016). Modeling the contribution of central versus peripheral vision in scene, object, and face recognition. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Wedel, M., & Pieters, R. (2015). The buffer effect: The role of color when advertising exposures are brief and blurred. *Marketing Science, 34*(1), 134–143.

Wei, X., Phung, S. L., & Bouzerdoum, A. (2016). Visual descriptors for scene categorization: Experimental evaluation. *Artificial Intelligence Review, 45*(3), 333–368.

Wu, C.-C., Wang, H.-C., & Pomplun, M. (2014). The roles of scene gist and spatial dependency among objects in the semantic guidance of attention in real-world scenes. *Vision Research, 105*, 10–20.

Wu, C.-T., Crouzet, S. M., Thorpe, S. J., & Fabre-Thorpe, M. (2015). At 120 msec you can spot the animal but you don't yet know it's a dog. *Journal of Cognitive Neuroscience, 27*(1), 141–149.

Yao, A. Y., & Einhäuser, W. (2008). Color aids late but not early stages of rapid natural scene recognition. *Journal of Vision, 8*(16), 12.

Scene gist processing and its mechanisms

LI Mengqi[1]; CHEN Zhimin[1]; ZHENG Yuanjie[2]; REN Yanju[1]

([1] *School of Psychology, Shandong Normal University, Jinan* 250358, *China*)

([2] *School of Information Science and Engineering, Shandong Normal University, Jinan* 250358, *China*)

Abstract: With a mere glimpse of a scene, observers can grasp a variety of perceptual and semantic information. This is referred to as scene gist. In recent years, scene gist processing has become an important topic in visual perception domain. Research on this theme can reveal the processing mechanisms for visual information and provide important implications for developing intelligent machine vision. The influencing factors, the controversial issues, and the neural basis of scene gist processing are reviewed. Some important issues, including the primary element of scene gist processing, the relevant theoretical explanations, the modulating factors of the hierarchical processing, the modulatory effect of attention, the timing dynamic characteristics and the construction of the brain function network, should be further explored.

Key words: scene gist; hierarchical processing; attention